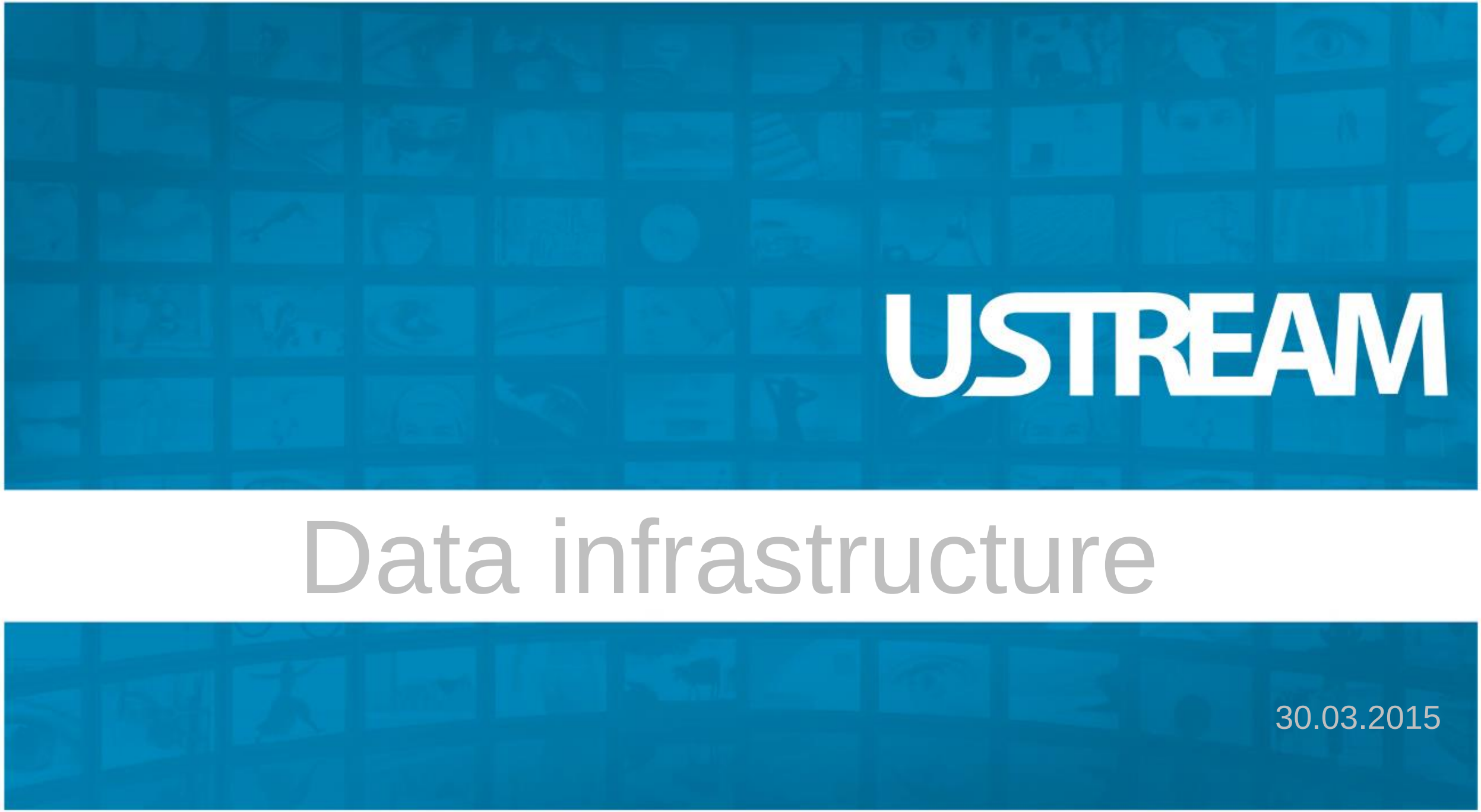


Data infrastructure

30.03.2015

Agenda

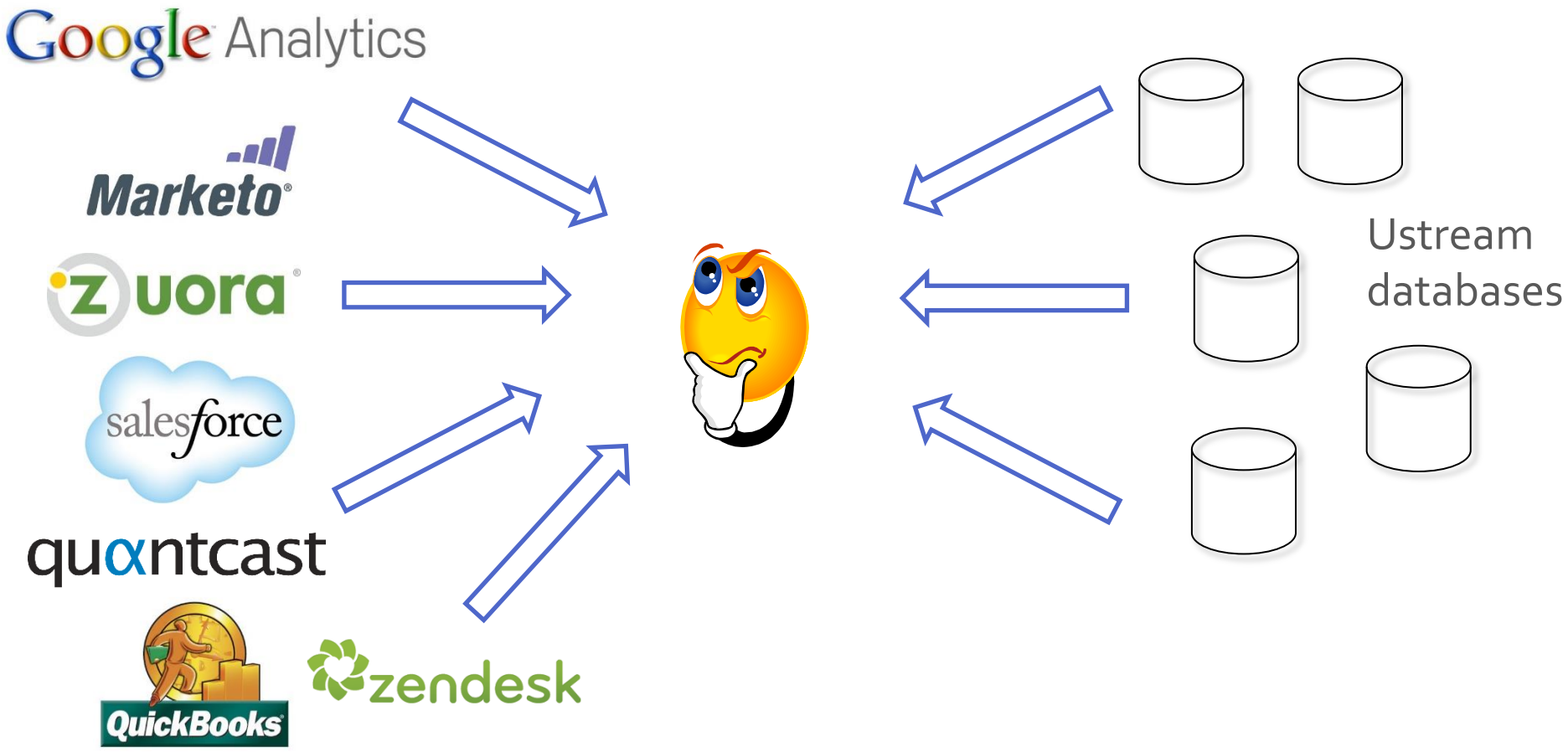
- A few words about Ustream
- Ustream data infrastructure before the BI team
- Current data infrastructure
- Big Data lessons learned & some future plans

Ever heard of Ustream?

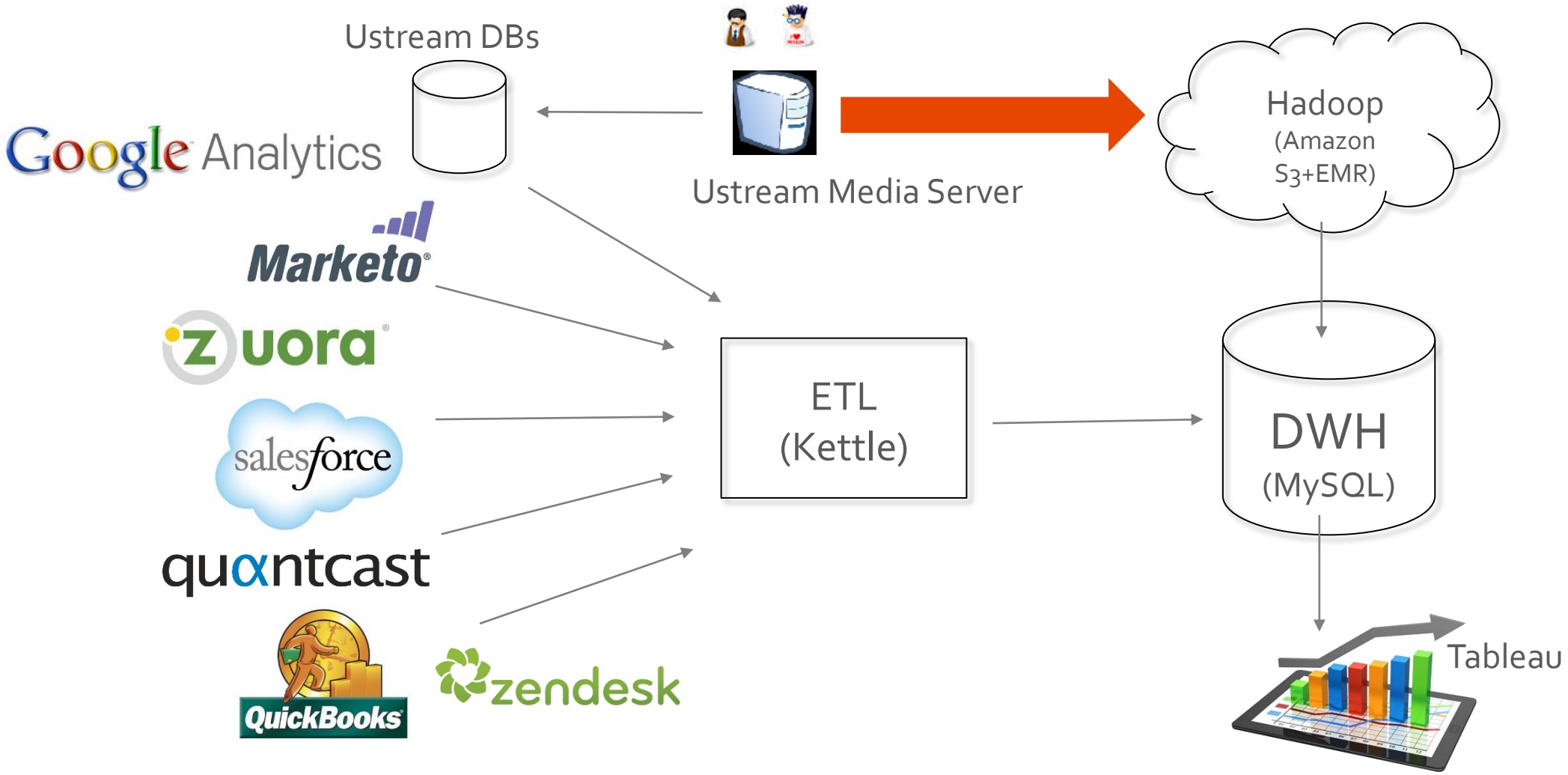
- World's leading live video service provider
- 80+ million monthly users
- SAAS company
- Founded in 2007
- 250 employees around the World
    - San Francisco, Budapest, Tokyo, and Seoul

Probably you've seen a few streams provided by us…

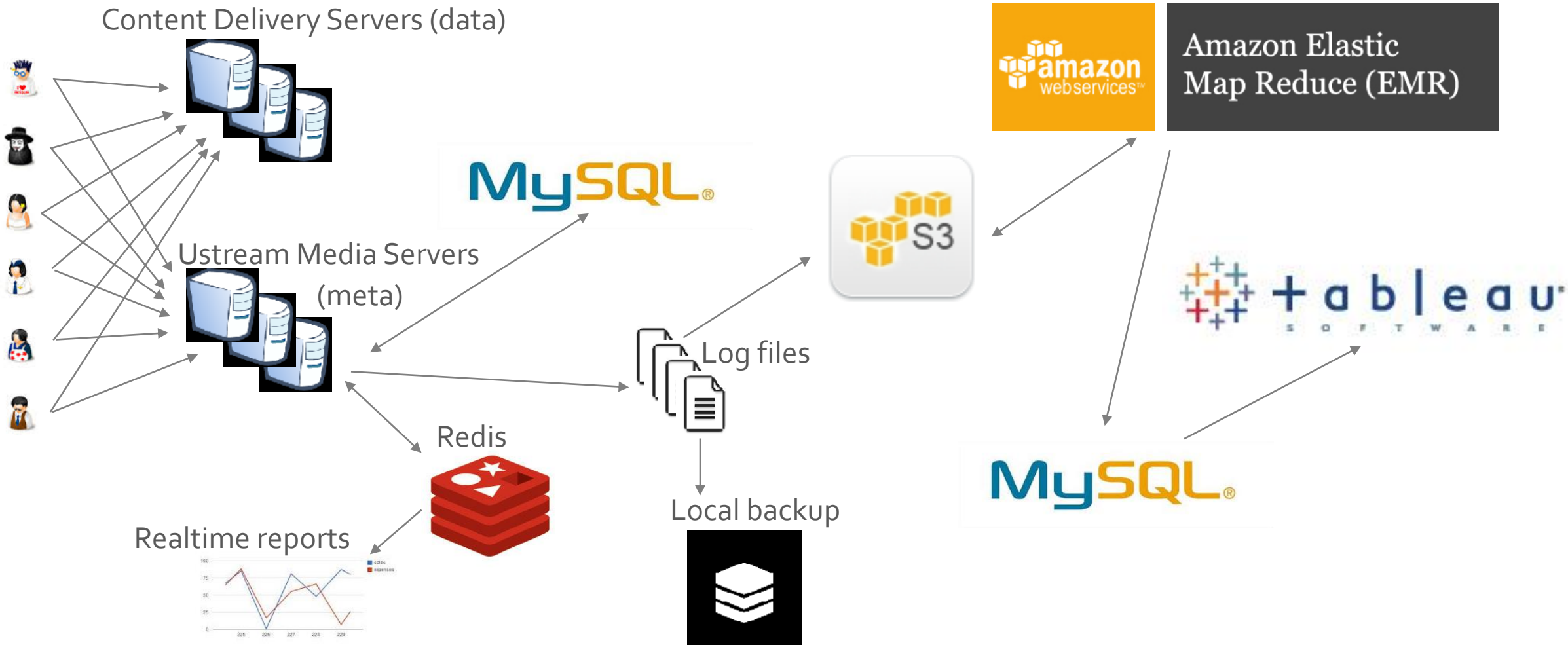
Data infrastructure – before BI team



Data infrastructure in general



Big Data infrastructure & data flow



Compromises in the architecture

- The architecture seems ad-hoc and heterogenous:
  - Yes, it is.
  - Important: no magic but still the problem is solved.
  - Fastest way to do the task with limited resources.

- It's a partial solution for Ustream's needs:
  - Financial, marketing, etc. reports go the traditional way
  - No reason to put small amount of data into Hadoop…

Lessons learned

- Big Data is not a buzzword for us anymore
  - Hadoop has some tricks but you can easily use it in production
  - Amazon EMR is a great place to learn
- Short time-to-market but with compromises
  - Small investment, still acceptable results
- Key factors:
  - strong sponsorship and trust from management
  - dedicated resources for research and development
  - user expectations had to be managed

Future plans

- Click-stream, buffering, usage pattern analysis
- Change in logging methods
  - Use Kafka for log shipping instead of log files
  - Merge logs into one to understand usage patterns
- Better self-serve interface needed
  - New version of Hue is promising
  - Tableau comes with direct Hadoop connector

Time to ask…

Q & A

Thanks for your attention!

szakacs.balazs@ustream.tv